Prediction of the growth in grower-finisher pigs using biologically constrained machine learning under small-sample data conditions

Yue Cheng[1], Yifei Tong[1], Honghua Huan[2*]

(1. *School of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China*;
2. *Jiangsu Academy of Agricultural Sciences, Nanjing 210014, China*)

Abstract: Accurate prediction of pig growth is essential for feed planning and market decisions in precision pig farming, but farm data are often small and fragmented. To address this challenge, a biologically constrained machine learning framework is proposed to predict the time required for grower-finisher pigs to reach the target weight of 100 kg under small sample conditions. By applying biologically constrained modifications to machine learning models (SC-GAM and Monotone XGBoost), the constrained models outperformed unconstrained baselines, which achieved higher accuracy (RMSE ≤ 4.73 d, ACC±7d ≥ 92%), greater robustness, and improved biological realism. An evaluation system was developed that combines traditional accuracy indicators with biologically grounded metrics. Practical applicability was examined via an on-farm shadow test on an independent batch. The models delivered reliable predictions that support finishing scheduling and feed-related management decisions. These findings highlight the potential of biologically constrained models to improve operational efficiency and reduce resource wastage in commercial pig farming.

Keywords: biologically constrained machine learning, pig growth prediction, small-sample data, precision livestock farming
DOI: 10.25165/j.ijabe.20261902.10243

Citation: Cheng Y, Tong Y F, Huan H H. Prediction of the growth in grower-finisher pigs using biologically constrained machine learning under small-sample data conditions. Int J Agric & Biol Eng, 2026; 19(2): 88–102.

1 Introduction

Precision livestock farming (PLF) is reshaping commercial pig production. By instrumenting animals and facilities, it generates continuous, high-frequency data streams to support management decisions. Recent reviews consistently identify growth monitoring and body weight attainment as high-value PLF targets because they directly affect feed planning, sorting and marketing windows, and welfare interventions[1]. Beyond periodic weighing, modern systems integrate scales and cameras with RFID tags, climate monitors, and feed-intake sensors to provide near-real-time views of herd dynamics. These data allow producers and integrators to coordinate logistics, adjust rations, and schedule shipments at scale[2,3]. As PLF moves from experimental barns into commercial practice, demand has grown for predictive tools that operate on routinely recorded traits, are robust to imperfect data capture, and deliver outputs that farm managers can interpret and act upon.

Research on PLF for pigs increasingly targets automated growth and body weight monitoring as core decision-support functions. Systematic reviews synthesize applications across sensing and analytics, while emphasizing that commercial deployments face fragmented data pipelines, heterogeneous devices, and often small per-farm sample sizes that complicate model development and external validation[4]. Vision-based systems and depth-based systems for live pig weight estimation, ranging from monocular and top-view imaging to RGB-D and point-cloud pipelines, achieve encouraging accuracy but remain sensitive to site conditions and dataset curation, underscoring the need for approaches that generalize under heterogeneous, small-sample regimes[5,6].

Existing modeling spans classical sigmoidal growth curves (e.g., logistic, Gompertz) and modern machine-learning regressors for tabular or multimodal inputs. At farm scale, tree ensembles and neural models have been widely adopted: deep networks trained on 3D point sets or depth maps can infer finishing weight from geometric description, while gradient-boosted trees remain strong baselines on structured traits owing to robustness and speed. However, unconstrained learners can produce predictions that violate biological regularities. This motivates shape-constrained additive models (SC-GAM), which encode monotonicity and curvature at the smoother level to deliver interpretable nonlinear effects under domain-informed restrictions[7]. In parallel, monotonic constraints in gradient boosting, such as XGBoost, enforce feature-wise orderings during split and leaf updates while retaining competitive accuracy and speed[8].

Given the prevalence of small, heterogeneous datasets in production barns, data augmentation serves as a complementary lever for data efficiency and variance control. Recent surveys catalog transformation-based, pattern-mixing, and generative strategies for time-series augmentation with updated taxonomies and head-to-head comparisons, offering practical recipes for forecasting and regression[9,10]. For tabular synthesis, diffusion-based models such as TabDDPM achieve state-of-the-art fidelity across public benchmarks and are openly released, facilitating adoption beyond vision tasks[11]. For rare or imbalanced targets, oversampling frameworks like SMOGN densify sparse outcome regions by combining interpolation and noise injection[12].

Evaluation methodology is equally critical for livestock applications. Methodological studies show that random *K*-fold cross-validation can yield optimistically biased error when windows from the same animal appear across folds. Grouped and blocked *K*-fold is

Received date: 2025-10-02 **Accepted date:** 2026-02-02
Biographies: **Yue Cheng**, MS candidate, research interest: industrial engineering, Email: yuecheng@njust.edu.cn; **Yifei Tong**, Professor, research interest: intelligent manufacturing, digitalized workshop management and control, Email: tyf1776@mail.njust.edu.cn.
***Corresponding author: Honghua Huan**, Deputy director, research interest: intelligent agriculture. Jiangsu Academy of Agricultural Sciences, Nanjing 210014, China. Tel: +86-15150515123, Email: hhhuan@jaas.ac.cn.

recommended for hierarchical or clustered data to avoid leakage and better emulate deployment[13]. In addition, there is growing emphasis on reporting calibrated uncertainty for regression via conformal prediction, including conformalized quantile regression, which offers finite-sample, distribution-free coverage and decision-relevant intervals[14,15].

Despite significant progress in PLF, current growth prediction models still face major challenges. One of the primary issues is the lack of biological constraints incorporated into both data augmentation and model design. Many models achieve high predictive accuracy yet ignore growth principles. For instance, they neglect the monotonic relationship between body weight and the time required to reach the target body weight, as well as the tapering of growth near target weight. As a result, these models may generate biologically unrealistic predictions, which are difficult to interpret and less useful for practical decision-making on farms.

Additionally, most existing models are not robust under small and heterogeneous datasets. Farm data is often noisy, incomplete, and fragmented. And per-farm sample sizes are small, which restricts generalizability. Traditional machine learning models struggle to maintain accuracy and interpretability when trained on such data, leading to overfitting and poor performance when deployed in real-world settings. Given these considerations, the need for biologically constrained evaluation approaches that account for both the biological realism and predictive accuracy of models becomes increasingly apparent.

To overcome such issues, this study introduces a biologically constrained machine learning framework that integrates both data augmentation and model design principles, ensuring that the predictions remain not only accurate but also biologically realistic. This framework offers a novel approach by respecting biological regularities, improving both prediction accuracy and model interpretability. This study makes the following key contributions:

1) Biologically Constrained Data Augmentation (BCDA): a new data augmentation method that generates biologically realistic synthetic data.

2) SC-GAM and Monotone XGBoost Models: two biologically constrained models enforcing monotonicity and capturing end-phase tapering for interpretable predictions.

3) Reverse Prediction Task: a novel formulation predicting the time to reach 100 kg instead of predicting future weight.

4) Integration of Evaluation Systems: a combination of traditional accuracy metrics and novel biologically constrained evaluation metrics.

5) Engineering Implementation: validation of the model's applicability through an on-farm shadow test, demonstrating its potential to support operational decision-making.

2 Materials and methods

2.1 Method outline

This study proposes a biologically constrained framework for predicting how long it would take for grower-finisher pigs to reach the target body weight of 100 kg, incorporating biologically constrained machine learning models (SC-GAM and Monotone XGBoost) and Biologically Constrained Data Augmentation (BCDA), as illustrated in Figure 1. The framework integrates rigorous data preprocessing, followed by biologically constrained data augmentation to enhance training datasets while maintaining biological realism. The predictive task is divided into early-stage and late-stage prediction tasks, focusing on the remaining time to reach 100 kg. Several machine learning algorithms are employed, with biologically constrained models outperforming unconstrained

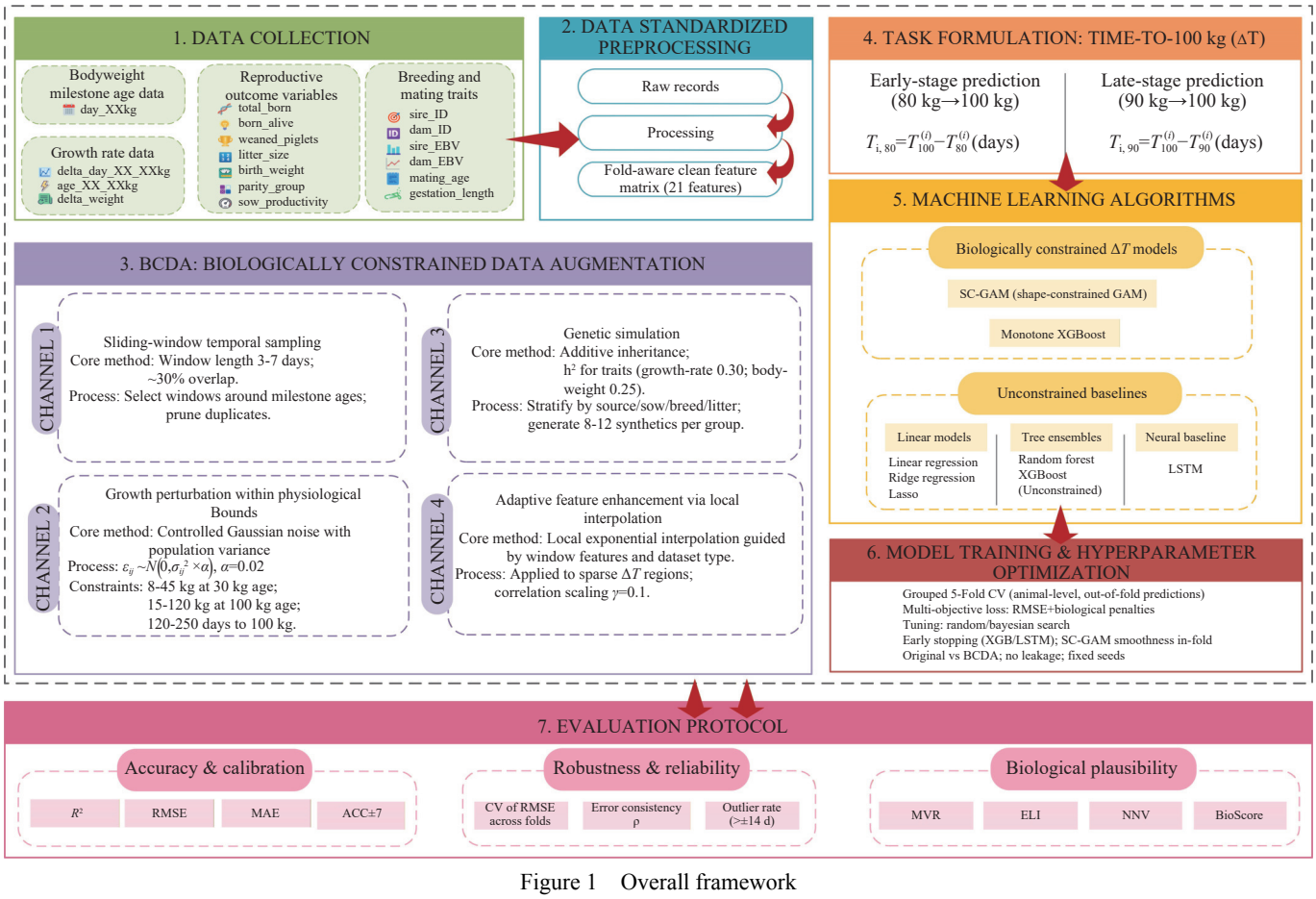


Figure 1 Overall framework

baselines, offering enhanced predictive accuracy and biological compliance. The models are evaluated using a combination of traditional accuracy metrics, robustness measures, and biologically grounded metrics to ensure both model performance and biological realism.

2.2 Data sources

The data for this study was sourced from J Farm in Taizhou, Jiangsu Province, China, which breeds Chinese native black pigs. The dataset covered the calendar year 2024 and included longitudinal growth records for the 30-100 kg stage. In total, 424 pigs from multiple batches were monitored.

Data included individual body weight records and litter-level reproductive and parental information, which were processed into features for modeling.

2.3 Variable definitions

The variables were organized into four categories for modeling: body weight milestone ages, growth-rate descriptors, reproductive and genetic records, and feeding and mixing traits. Table 1 summarizes definitions and units, with full descriptions provided in Table 2.

Table 1 Grouped variables used in modeling

No.	Variable type	Examples
1	Body weight milestone ages	day_30kg, day_40kg, day_50kg, day_60kg, day_70kg, day_80kg, day_90kg, day_100kg
2	Growth-rate descriptors	delta_day_30_60, delta_day_30_90, delta_day_30_100, delta_weight_30_60, delta_weight_30_90, delta_weight_30_100, adg_30_60, adg_30_100
3	Reproductive and maternal records	total_born, born_alive, stillborn, mummified, deformity, weaned_piglets, birth_weight_live_weak, avg_birth_weight, avg_weaning_weight, parity_group, litter_size_group, sow_productivity, EBV_difference
4	Breeding and genetic records	sire_ID, sire_mating_age, sire_EBV, dam_ID, dam_mating_age, dam_EBV, parity, gestation_length

Table 2 Full descriptions used in modeling

No.	Variable name	Description
A1	Age/d	Days to reach target weight
A2	Delta day	Days between 30 and 60 kg, 30 and 90 kg, 30 and 100 kg
A3	Delta weight/kg	Weight gain from 30 to 60 kg, 30 to 90 kg, 30 to 100 kg
A4	ADG	Average daily gain
A5	Total born	Total piglets born per litter
A6	Born alive	Number of live-born piglets
A7	Stillborn	Number of stillbirths
A8	Mummified	Number of mummified fetuses
A9	Deformity	Number of deformed piglets
A10	Weaned piglets	Number of piglets weaned
A11	Birth weight live weak/kg	Total birth weight of weak live piglets
A12	Avg birth weight/kg	Average birth weight per litter
A13	Avg weaning weight/kg	Average weaning weight
A14	Grouping by parity	Parity classification of the dam
A15	Grouping by litter size	Classification based on total number of piglets born
A16	Sow productivity	Metrics like live births per parity, weaning rate, etc.
A17	Sire's age at first mating	Days of age when sire first used for breeding
A18	Sire's EBV	Estimated Breeding Value of the sire
A19	Dam's age at first mating	Days of age when dam first used for breeding
A20	Dam's EBV	Estimated Breeding Value of the dam
A21	Parity of the dam	Parity number of the sow at farrowing
A22	Gestation length	Gestation period in days
A23	EBV difference (Sire-Dam)	Difference in breeding values between sire and dam
A24	Ear mark, litter no	Individual pig ID and litter number

Beyond raw fields, several derived variables were computed to enhance model representation. These included:

1) Parity group: Parity was classified to control for its effects on offspring performance and growth;

2) Litter-size group: Samples were grouped into small, medium, and large based on litter size distribution;

3) Sow productivity: Reproductive efficiency was quantified using ratios such as survival rate excluding stillbirths and weaned piglets to live births ratio;

4) EBV difference (Estimated Breeding Value Difference): Calculated from sire and dam breeding values to capture genetic potential differences affecting growth performance.

All variables were standardized and used as input features for model training. As listed in Table 3, the dataset abstracts growth curves into discrete data points, with Figure 2 visualizing age distribution at each milestone and highlighting growth rate fluctuations for predictive modeling.

Table 3 Statistical summary of the dataset

	Measurement days	Days to 30 kg	Days to 100 kg	Boar breeding value	Sow breeding value
5th percentile	16.00	75.05	172.35	65.23	59.33
Lower quartile	48.00	88.00	187.00	92.97	83.21
Mean	67.00	97.00	196.00	111.44	97.37
Median	64.37	96.32	197.15	117.82	97.55
Upper quartile	79.50	105.00	206.00	154.59	114.26
95th percentile	108.00	115.30	225.00	165.27	132.50
Standard deviation	26.04	12.70	15.58	33.44	23.70

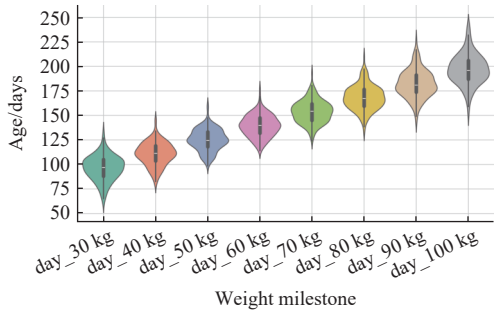


Figure 2 Violin plot of age at bodyweight milestones

2.4 Data preprocessing

Age at each body weight milestone is expected to follow a smooth, biologically realistic trajectory. This study therefore adopted a Median Absolute Deviation (MAD) to identify records that deviated significantly from normal developmental patterns. This approach is more robust to extreme fluctuations than conventional statistical thresholds, making it particularly suitable for detecting anomalies caused by recording errors or biological outliers.

The anomaly detection pipeline was applied to each individual pig's body weight-age sequence as follows:

1) Remove all body weight milestone records below 30 kg (as the starting weights of all pigs exceed this threshold).

2) Sort each pig's body weight milestones from smallest to

largest in sequential order.

3) Fit a quadratic regression to the observed milestone pairs.

$$\hat{t}(w) = \beta_0 + \beta_1 w + \beta_2 w^2 \quad (1)$$

4) Perform quadratic polynomial regression analysis on the sorted age values against their corresponding body weight values.

5) Use the regression model to predict the theoretical age for each individual body weight milestone.

6) Calculate the median and median absolute deviation (MAD) for all deviations.

7) Separately calculate the median and median absolute deviation (MAD) for all deviations.

8) Establish upper and lower boundaries based on the following formula to distinguish between normal and abnormal values:

$$\tau = \mathrm{Med}(e) + \lambda \mathrm{MAD}(e) \quad (2)$$

9) Mark records where meets the requirements of Equation (3).

$$e_j > \tau \quad (3)$$

10) Remove records identified as outliers through deletion processing.

To improve anomaly detection, λ was treated as a tunable hyperparameter, adjusted based on dataset characteristics. Fixed thresholds (e.g., λ = 1.0 or 2.5) may be inadequate for pig growth data due to its high volatility and low anomaly rates. Therefore, λ values were selected to balance outlier removal and model stability, with λ = 2.5 chosen after sensitivity testing. Detailed diagnostics are listed in Table 4.

Table 4 Effect of parameter λ on outlier removal performance

λ value	Number of removed samples	Number of retained samples	Removal rate/%	Average boundary width
1.5	180	1164	13.39%	48.6
2.0	150	1194	11.17%	56.7
2.5	127	1217	9.45%	65.1

To obtain a consistent modeling matrix, this study applied a standardized preprocessing pipeline to all variables. Figure 3 illustrates the data process flow used in this study, showing the sequence of steps for outlier detection, feature engineering, and the final feature matrix construction.

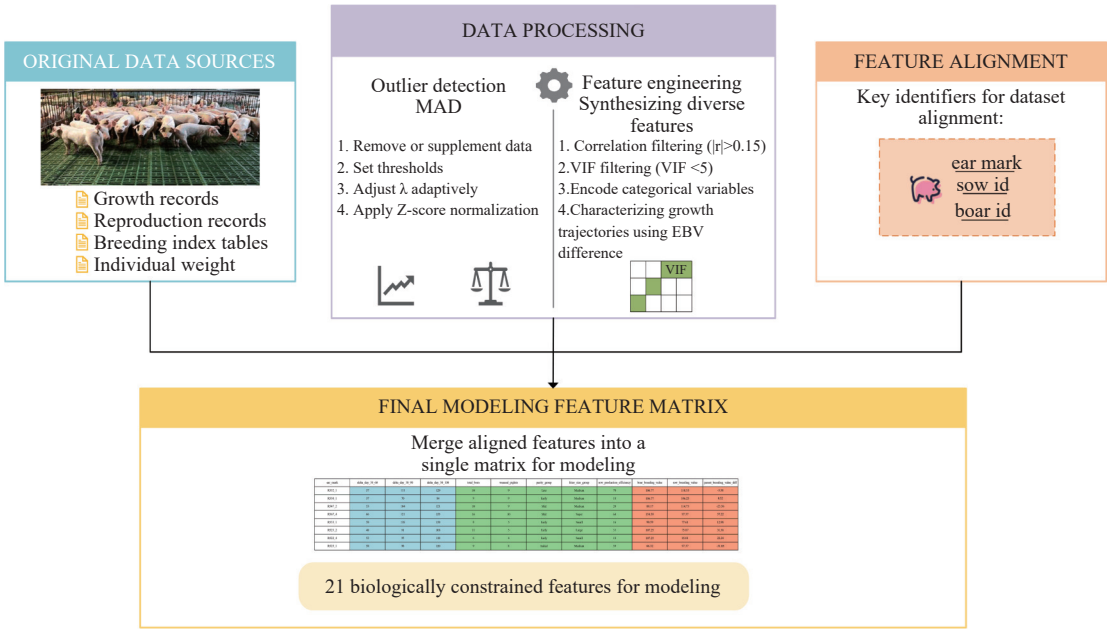


Figure 3 Data process flow

2.5 Dataset generation with biologically constrained data augmentation

This section presents the Biologically Constrained Data Augmentation (BCDA) framework for small-sample regimes. Figure 4 illustrates BCDA process flow used in this study. BCDA increases training instances while preserving data validity and strict splitting, consisting of four components: temporal sampling with biological constraints, growth perturbation within biological limits, genetic simulation with biological considerations, and feature enhancement based on biological patterns. In practical farming, limited batch sizes, cost constraints, and incomplete records often yield small samples and distribution shift, leading to unstable model selection and inflated evaluation variance; therefore, all augmented instances carry a synthetic flag $z \in \{0,1\}$ (1 = synthetic) and are kept within the same cross-validation fold as their source individual to prevent leakage.

2.5.1 Temporal sampling with biological constraints

The sliding window technique is the core method for processing pig growth temporal data, capturing dynamic features across different growth stages[16,17]. Each individual i has an ordered weight-age series $\{(w_{it}, t_{it})\}_{t=1}^{T_i}$. A window of length L and step s moves along the series to form subsequences.

$$W_{i,l} = \{(w_{it},\ t_{it})\}_{t=l}^{l+L-1},\ l = 1, 1+s, \ldots, T_i - L + 1 \quad (4)$$

L is specified either in days or observations (e.g., window lengths of 3-7 d with ~30% overlap). Each $W_{i,l}$ is summarized into $x_{i,l}$ using the taxonomy in Section 1.3. This study adopts a selective strategy that targets windows covering key body weight milestones (e.g., 30, 60, 90, 100 kg) and then prunes near-duplicate windows to avoid excessive augmentation.

2.5.2 Growth perturbation within biological bounds

To introduce controlled stochastic variation while preserving biological constraints, a perturbation was applied to observed weights or ages within each window, following principles of biological variation modeling[18]. At time point j for individual i, the perturbed value is shown below.

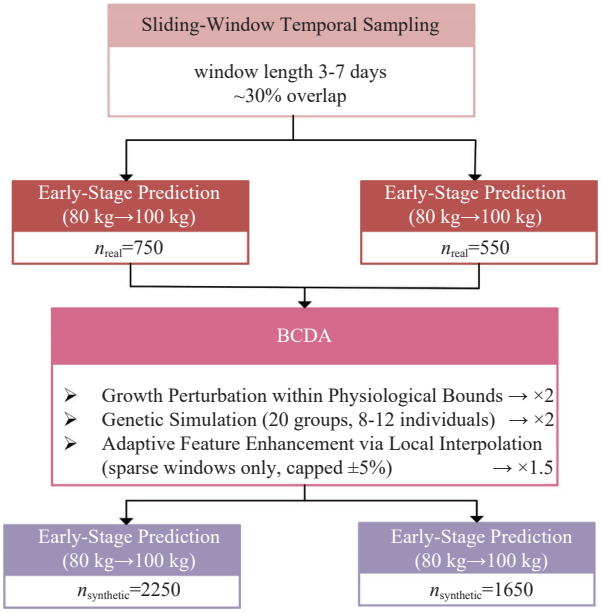


Figure 4 BCDA process flow

$$x'_{ij} = x_{ij} + \varepsilon_{ij}, \varepsilon_{ij} \sim N\left(0, \sigma_{ij}^2 \times \alpha\right), \quad (5)$$

where, x_{ij} denotes the original measurement (weight or milestone age, depending on the descriptor), σ_{ij}^2 is the variance estimate for that feature at that sample, and α is the noise-scaling factor. In this study, σ_{ij}^2 is estimated within each cross-validation fold to prevent information leakage.

To assess the robustness, three noise-scaling factors ($\alpha \in \{0.01, 0.02, 0.04\}$) were tested. For each value of α, perturbed samples are regenerated and all models are retrained under identical hyperparameter settings. Model performance is then evaluated using RMSE, MAE, and ACC±7d, with the corresponding sensitivity curves reported in Figure 5. Based on this analysis, α=0.02 is chosen for subsequent experiments.

After perturbation, all derived descriptors were recalculated. This study then enforced biologically constrained growth by ensuring monotonic increases in weight and age, with realistic daily gains (e.g., ADG ∈ [0.3, 1.2] kg/d)[19], and finishing ages consistent with the literature (e.g., about 120-250 d to reach 100 kg)[20]. Any violation triggers rejection and resampling.

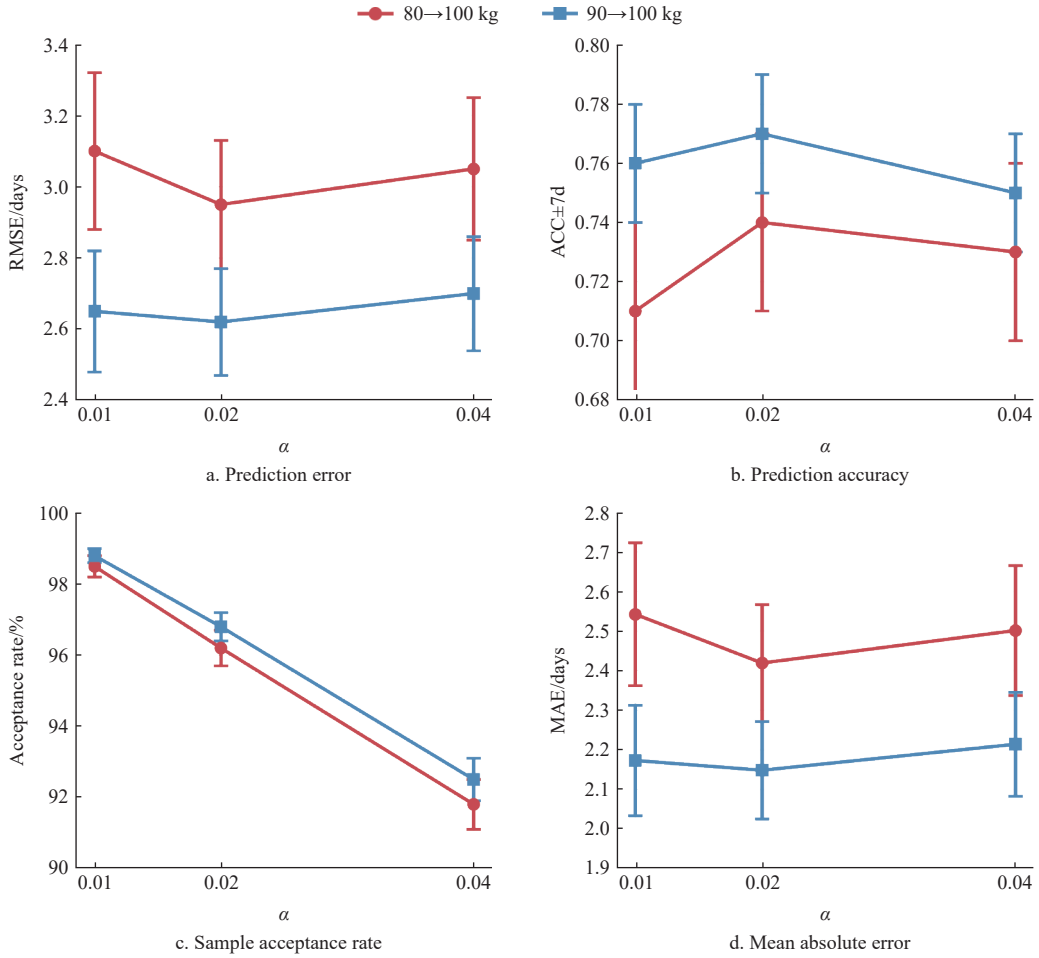


Figure 5 Sensitivity of model performance to noise-scaling factor α

2.5.3 Genetic simulation with biological considerations

Grounded in quantitative genetics, this study generated synthetic individuals within existing genetic groups using an additive-inheritance formulation[21,22]. Trait-specific heritability coefficients are adopted (e.g., $h^2 \approx 0.30$ for growth-rate traits; $h^2 \approx 0.25$ for body weight traits)[23]. Let the phenotype be

$$Y_{ij}(t) = \mu_j(t) + G_{ij}(t) + E_{ij}(t), \quad (6)$$

where $\mu_j(t)$ denotes the time-dependent population mean, $G_{ij}(t) \sim N(0, \sigma_a^2)$represents the additive genetic effect, and $E_{ij}(t) \sim N(0, \sigma_e^2)$ represents the environmental effect. The heritability is $h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$. To retain structure and realism, this study stratifies individuals into 20 genetic groups using sire ID, sow ID, breed, and litter background; within each group, 8-12 synthetic individuals are generated while preserving intra-group consistency and allowing

inter-group diversity. The simulated genetic offsets are applied to body-rate descriptors and EBV-related proxies, and are capped so that implied ADG and milestone ages remain within fold-wise empirical bounds.

2.5.4 Feature enhancement based on biological patterns

For sparsely represented segments such as long intervals between weightings, this study adopted a local exponential and linear interpolation strategy that adapted to window features and dataset type[24]. The enhanced trajectory is

$$W'_{\text{enhanced}}(t) = W(t) \times \left(1 + \gamma \times f\left(\text{window}, \text{dataset\_type}\right)\right) \quad (7)$$

which subjects to shape constraints (non-decreasing weight; $\Delta w/\Delta t$ within a realistic ADG range). This study implemented dataset-specific heuristics: breeding-performance datasets adjust growth parameters according to reproductive efficiency indices; growth-increment datasets adjusted rates by age increments; baseline datasets emphasize growth-curve consistency; breeding-group datasets combine paternal and maternal genetic information. For a small subset per dataset, this study additionally interpolated using a local exponential form:

$$W(t) = W_0 \times \exp\left(\beta(\text{window}) \times \left(1 - \exp\left(-\gamma \cdot (\text{window} \times t)\right)\right)\right) \quad (8)$$

And this study capped the final adjustment at ±5% to maintain plausibility and correlation structure. This procedure typically yields 1-2 enhanced variants per qualified window.

2.6 Task modeling: time-to-100 kg (ΔT)

In the national and industry guidelines, 100 kg body weight is widely used as the threshold for slaughter or market readiness. To support precision pig-farming decisions, this study predicts the remaining days for an individual to reach 100 kg, providing a directly actionable timeline. Forward prediction of future weight at fixed horizons (e.g., 7, 14, 21 d) can be ambiguous, as the output is a weight value rather than a timeline. To align with on-farm decisions, this study adopts a reverse formulation, predicting the remaining time to reach 100 kg in days.

For individual i, let $T_{100}^{(i)}$ denote the day the pig first reaches 100 kg (milestone age) and let $T_{\theta}^{(i)}$ denote the day it first reaches the threshold $\theta \in \{80, 90\}$ kg. The task label is

$$\Delta T_{i,\theta} = T_{100}^{(i)} - T_{\theta}^{(i)}(d) \quad (9)$$

The early-stage task uses θ=80 kg for decisions 20-30 d before finishing, while the late-stage task uses θ=90 for actions 10-15 d before slaughter.

Let $\{(w_{it}, t_{it})\}_{t=1}^{T_i}$ be the ordered body weight-age series for pig i over 30-100 kg. From each trajectory this study obtained windows $W_{i,l}$ by the sliding protocol in Section 2.5.1. A window produces a supervised sample only if it covers the relevant threshold:

$$S_{80\to100} = \left\{ \left(x_{i,l}, y_{i,l}\right) : W_{i,l} \text{including 80 kg},\ y_{i,l} = \Delta T_{i,80} \right\} \quad (10)$$

$$S_{90\to100} = \left\{ \left(x_{i,l}, y_{i,l}\right) : W_{i,l} \text{including 90 kg},\ y_{i,l} = \Delta T_{i,90} \right\} \quad (11)$$

where, $x_{i,l}$ is the feature vector aggregated from body weight milestone ages, rate-based descriptors, reproductive and maternal records, and breeding and genetic records. BCDA-generated samples inherit the source individual's identifiers and remain in the same cross-validation fold.

For model stability, ΔT is handled as an integer-valued outcome in days. If $\Delta T = 0$ arises from same-day threshold and finishing events, it is retained as a valid label. All time stamps are computed from the earliest date a threshold is first reached to avoid double counting in fluctuating measurements. The next section specifies the modeling principles, biological constraints, and multi-objective optimization used to learn predictors of ΔT under small-sample conditions.

3 Research methods

3.1 Biologically constrained ΔT model and unconstrained baselines

The prediction target is $y = \Delta T$, representing the remaining number of days for an individual to reach 100 kg from a task threshold $\theta \in \{80, 90\}$ kg. For pig i, let $T_{100}^{(i)}$ be the first day it reaches 100 kg and $T_{\theta}^{(i)}$ the first day it reaches θ. Predictors x are constructed strictly from available information to prevent leakage, and all windows originating from the same individual remain in the same cross-validation fold thereafter.

To integrate biological constraints into predictive modeling, this study used two complementary models: SC-GAM and Monotone XGBoost. These models incorporate biological priors to ensure results are biologically realistic and predictive.

SC-GAM is a shape-constrained generalized additive model. It incorporates monotone P-splines and localized curvature control to enforce biological constraints. The predicted time decreases with higher body weight ($\partial f/\partial w \le 0$), increases with higher ADG ($\partial f/\partial \text{ADG} \ge 0$), and shows a gradually decreasing trend as the finishing weight is approached. Monotonicity is enforced through non-negative differences between adjacent spline coefficients, and tapering is induced by a curvature penalty within the 90-100 kg range. Additionally, predictions are constrained to be non-negative through a hinge penalty.

Monotone XGBoost leverages a gradient-boosting framework that allows for complex feature interactions while maintaining monotonicity. This model uses a feature-wise monotonicity vector $m \in \{-1, 0, 1\}^P$, where -1 corresponds to weight variables, $+1$ corresponds to ADG-type rates, and 0 corresponds to categorical groups. Interaction constraints are applied to ensure global monotonicity for designated features, ensuring that the model respects biological growth principles. The early- and late-stage tasks are fitted separately and coupled during training by a consistency term, ensuring that early-task predictions are always greater than or equal to those for late-stage tasks, reflecting biological progression.

Monotonicity and interaction constraints are disabled. The following unconstrained models are trained on the same ΔT tasks: Linear Regression, Ridge, Lasso, Elastic-Net, Single-tree CART, Random Forest, Unconstrained XGBoost, and LSTM. All these models share the same feature boundaries and fold assignments as the biologically constrained models, providing transparent linear references and non-monotone tree-based comparisons for assessing predictive accuracy, stability, and biological realism.

3.2 Training

All experiments were conducted using Python 3.9.10 on a desktop computer with Intel Core i5-13600KF (14 cores, 3.50 GHz), 32 GB RAM, and NVIDIA GeForce RTX 4060 Ti (16 GB VRAM).

Monotone XGBoost models were configured with 1000 estimators, learning rate 0.05, and maximum depth 6. Regularization was applied using L_1 (α=0.1) and L_2 (λ=1.0) penalties to prevent overfitting while capturing complex relationships. Monotone constraints were applied to biologically relevant features, such as weight milestones and growth rates, to align predictions with growth patterns.

SC-GAM models used 6-10 knot penalized splines fitted via REML estimation in R with the MGCV package, ensuring smooth, biologically plausible growth patterns. Conformal prediction was applied with 90% and 95% coverage for reliable uncertainty estimates.

Models were trained with five-fold cross-validation, partitioned by pig ID to prevent data leakage. GPU acceleration was used for XGBoost training, reducing time to approximately eight minutes per fold. Early stopping was applied based on validation RMSE with a 50-round patience.

3.3 Metrics and tests

To evaluate the model performance, this study employed a combination of traditional accuracy metrics and novel biologically constrained evaluation metrics.

For the traditional accuracy metrics, this study used the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2). To evaluate the model's ability to predict within an acceptable tolerance window, this study introduced Tolerance Accuracy (ACC±7d). In addition, the CV of RMSE was computed to assess fold-to-fold consistency.

$$\mathrm{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}, \tag{12}$$

$$\mathrm{MAE}=\frac{1}{n}\sum_{i=1}^{n}|y_i-\hat{y}_i|, \tag{13}$$

$$R^2=1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}, \tag{14}$$

$$\mathrm{ACC}_{\pm k}=\frac{1}{n}\sum_{i=1}^{n}1\{|y_i-\hat{y}_i|\le k\},\ k=7, \tag{15}$$

$$\mathrm{CV}_{\mathrm{h}}=\frac{\mathrm{sd}(\mathrm{RMSE}_{\mathrm{f,h}})}{\mathrm{mean}(\mathrm{RMSE}_{\mathrm{f,h}})} \tag{16}$$

In addition to traditional accuracy metrics, this study introduced biologically grounded metrics to assess the biological plausibility of model predictions, ensuring predictions follow basic biological principles. The evaluation included: Monotonicity Violation Rate (MVR), Early-Late Consistency Index (ELI), and Non-Negativity Violations (NNV). MVR assessed the monotonic relationship between body weight and time-to-finish, ELI ensured early-stage predictions were greater than or equal to late-stage predictions, and NNV prevented negative time-to-finish values.

$$\mathrm{MVR}=\frac{\sum_{i}\sum_{j}1(\hat{y}_i(w_{j+1})-\hat{y}_i(w_j)>0)}{\sum_{i}(n_i-1)} \tag{17}$$

$$\mathrm{ELI}=1-\frac{\sum_{i}\max(0,\hat{y}_i^{(90)}-\hat{y}_i^{(80)})}{\sum_{i}\hat{y}_i^{(80)}}\in[0.1] \tag{18}$$

$$\mathrm{NNV}=\frac{\sum_{i}1(\hat{y}_i<0)}{n} \tag{19}$$

These components were combined into a Biological Score, which evaluates the overall biological realism of the model's predictions:

$$\mathrm{BioScore}=\alpha(1-\mathrm{MVR})+\beta\mathrm{ELI}+\gamma(1-\mathrm{NNV}) \tag{20}$$

In addition to these metrics, this study calculated the Outlier Rate, which is defined as the proportion of samples whose prediction error exceeds 14 d.

$$\mathrm{OutlierRate}=\frac{1}{n}\sum_{i=1}^{n}1(|y_i-\hat{y}_i|>14\ \mathrm{d}) \tag{21}$$

To quantify the uncertainty in the predictions, this study used grouped bootstrap resampling (1000 resamples) and reported 95% confidence intervals for RMSE, MAE, R^2, and ACC±7d, providing a robust estimate of model variability.

For global performance comparisons, this study used the Friedman test on mean ranks across the N blocks. Additionally, Kendall's effect size τ was used to quantify the strength of performance differences. Post-hoc comparisons were performed using the Nemenyi test at α=0.05. Pairwise comparisons were confirmed using two-sided Wilcoxon signed-rank tests. The notations mentioned in the equations are listed in Table 5.

Table 5 Notation list

Variables (Observed/ Predicted)	Description
y_i	True label value for the i-th sample
$\hat{y}_i$	Predicted value for the i-th sample
$\bar{y}$	Mean of the true labels across all samples
w_j	Weight value for the j-th sample
n_i	Number of data points associated with the i-th sample used for calculating MVR
$\hat{y}_i^{(90)}$	Predicted weight for the i-th sample at 90 kg milestone
$\hat{y}_i^{(80)}$	Predicted weight for the i-th sample at 80 kg milestone
Evaluation metrics	
$\mathrm{RMSE}_{f,h}$	Root Mean Squared Error (RMSE) for fold f and horizon h
Functions/operators	
$1(\cdot)$	Indicator function that returns 1 if the condition is true, otherwise 0
Parameters α,β,γ	Weight parameter for the composite Biological Score, typically (α,β,γ) = (0.30,0.40,0.30)

4 Results

4.1 Data consistency analysis

To evaluate the quality of synthetic data generated by the BCDA framework, KS tests and QQ plots were used to compare synthetic and real samples. The KS test was performed on key variables to assess distributional consistency between real and synthetic data. The results are summarized in Table 6.

Table 6 Statistical comparison of real and synthetic samples

Variable	KS_stat	p_value	Cohen_d
day_30kg	0.0698	0.7812	0.022
day_60kg	0.0889	0.4123	0.053
day_80kg	0.0934	0.3456	0.048
day_90kg	0.0823	0.5123	0.045
day_100kg	0.0956	0.2856	0.062
delta_day_30_90	0.0712	0.7234	0.038
delta_day_30_100	0.1045	0.1567	0.071

The KS test shows that for most variables, synthetic and real data distributions are statistically similar, confirming the reliability

of BCDA-generated data. However, for delta_day_30_100, the *p*-value is 0.1567, indicating a small discrepancy that may require further refinement in the data generation process.

The QQ plots for key variables, including day_90kg, day_100kg, and delta_day_30_100, are presented in Figure 6. The plots revealed that the synthetic data points align closely with the real data points along the reference line, suggesting that the distributions of the real and synthetic samples were similar.

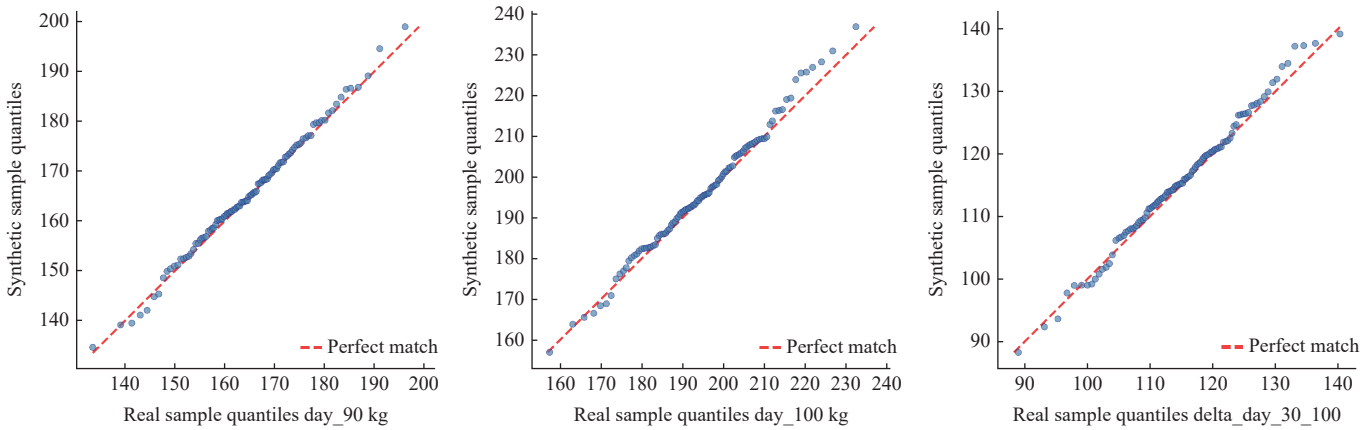


Figure 6 QQ plots comparing the distribution of real and synthetic samples

4.2 Overall predictive performance on Δ*T* tasks

This study investigated the early-stage task $\Delta T_{80\to100}=T_{100}-T_{80}$ and the late-stage task $\Delta T_{90\to100}=T_{100}-T_{90}$ with all models trained on data up to the corresponding threshold θ and BCDA-augmented datasets. Grouped cross-validation was used to prevent cross-individual leakage.

Biologically constrained models outperformed others in both tasks (Tables 7 and 8). As the remaining time shortens from $\Delta T_{80\to100}$ to $\Delta T_{90\to100}$, absolute errors decrease, but constrained models show a relative performance gain near finishing. In particular, SC-GAM improves RMSE and MAE over the best unconstrained baseline by about 20% and 10% to 15% on $\Delta T_{90\to100}$. Constrained models also deliver the highest R^2 and the largest share of predictions within an actionable window. These patterns are consistent with the design: monotonicity and tapering constraints regularize end-phase, namely the dynamic changes of approximately 90 to 100 kg, preventing biologically unrealistic overshoot and stabilizing short-horizon forecasts.

Table 7 Model performance comparison for early Δ*T* prediction task

Model		RMSE	MAE	R^2	ACC±7d
Biology-constrained models	SC-GAM	8.42	6.18	0.847	0.892
	Monotone XGBoost	9.15	6.84	0.824	0.875
Unconstrained baselines	Linear regression	12.36	9.45	0.672	0.743
	Ridge regression	11.84	8.97	0.693	0.761
	Lasso	12.08	9.12	0.685	0.754
	ElasticNet	11.97	9.03	0.688	0.758
	Random forest	9.87	7.34	0.789	0.842
	XGBoost (unconstrained)	9.63	7.15	0.798	0.849
	LSTM	10.42	7.89	0.761	0.825

The widening advantage of biology-constrained learners in the late horizon is informative. As the time window shortens, errors stem more from end-phase variability, particularly the weight increase from 90 to 100 kg. SC-GAM's tapering penalty explicitly regularizes this region, and its shape constraints prevent biologically unrealistic reversals when current weight or recent gain is high; monotone XGBoost achieves a similar effect through feature-wise monotonic splits. In contrast, unconstrained models can trade a small gain in in-sample fit for higher tail risk, which is not rewarded by ACC±7d.

Table 8 Model performance comparison for late Δ*T* prediction task

Model		RMSE	MAE	R^2	ACC±7d
Biology-constrained models	SC-GAM	4.73	3.45	0.823	0.924
	Monotone XGBoost	5.21	3.82	0.801	0.908
Unconstrained baselines	Linear regression	7.89	5.92	0.358	0.785
	Ridge regression	7.52	5.68	0.637	0.798
	Lasso	7.71	5.81	0.664	0.791
	ElasticNet	7.63	5.74	0.669	0.794
	Random forest	5.94	4.28	0.751	0.876
	XGBoost (unconstrained)	5.87	4.21	0.758	0.872
	LSTM	6.35	4.67	0.716	0.844

4.3 Statistical significance and rank-based comparison

To verify that the observed performance gains are not due to chance, this study conducted a global, rank-based comparison across both Δ*T* tasks and all accuracy metrics, including RMSE, MAE, R^2, and ACC±7d. For each block defined by {horizon×metric×CV-fold}, algorithms were ranked. A lower ranking means better performance, and the mean ranks were analyzed with the Friedman test followed by Nemenyi post-hoc comparisons. The Friedman statistic strongly rejects the null hypothesis of equal mean ranks (χ^2=47.82, p<0.001), establishing overall differences among learners.

The biology-constrained family occupies the top of the league table. SC-GAM attains the best mean rank, which is 1.23, and Monotone XGBoost follows closely at 1.89; both belong to the highest significance group A, clearly separated from the unconstrained baselines. The strongest baseline, unconstrained XGBoost, with a mean rank of 3.45, and Random Forest, with a mean rank of 3.78, form group B. Neural and linear baselines are consistently worse, falling into groups C-D (Table 9, Figure 7).

The rank-based analysis shows that improvements are not tied to any single metric or horizon. Practically, this indicates robustness of the constrained family under different operational utilities. The separation observed in the Nemenyi groups suggests that adding shape and monotonicity constraints changes the error geometry rather than only shifting the bias-variance trade-off.

4.4 Ablation studies

4.4.1 Ablation on BCDA

This subsection evaluates the contribution of the proposed

BCDA on both early-stage and late-stage prediction tasks by comparing three data settings: No BCDA, Temporal sampling only, and Full BCDA.

Table 10 and Table 11 summarize the results. In both tasks, Temporal-only outperformed the baseline, indicating that sliding-window sampling increases effective supervised instances. Full BCDA further improved all metrics compared with Temporal-only, suggesting that the additional gains were driven by biologically constrained synthesis rather than sample-size expansion alone. The improvement was more pronounced in early-stage prediction, consistent with its higher data sparsity and longer prediction horizon.

Table 9 Model ranking by Friedman test with Nemenyi post-hoc (A best)

Rank	Model	Mean rank	Significance group
1	SC-GAM	1.23	*A*
2	Monotone XGBoost	1.89	*A*
3	XGBoost (unconstrained)	3.45	*B*
4	Random forest	3.78	*B*
5	LSTM	4.92	*C*
6	Ridge regression	5.34	*C*
7	ElasticNet	5.67	*C*
8	Lasso	5.89	*C*
9	Linear regression	6.83	*D*

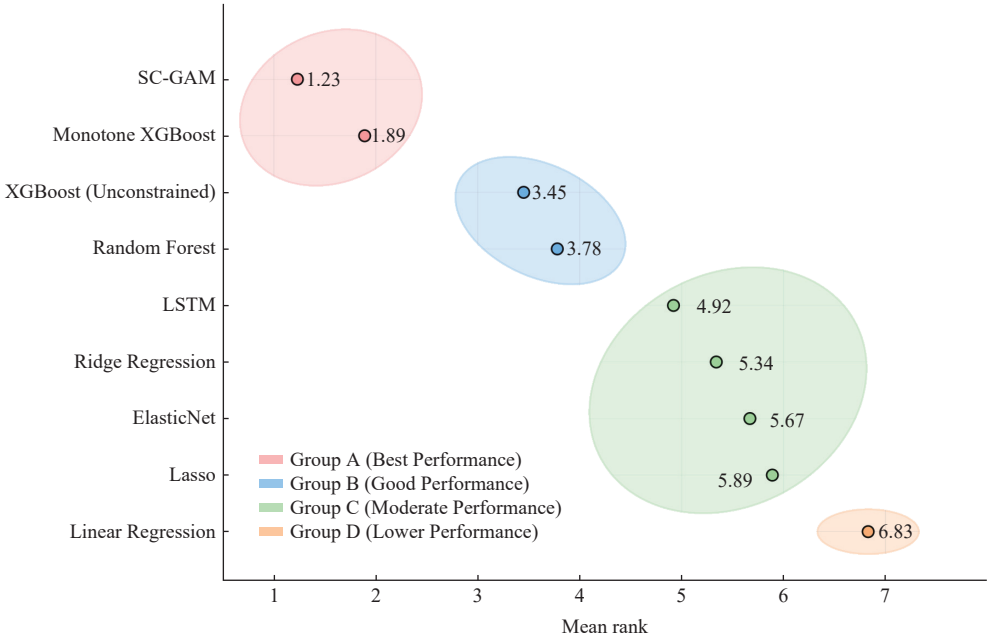


Figure 7 Rank-based comparison (Friedman and Nemenyi)

Table 10 Ablation results under different data augmentation strategies for early-stage prediction

Method	MAE	RMSE	R^2	MAPE	ACC±7d
No BCDA	9.45	12.36	0.672	5.38	0.743
Temporal sampling only	7.15	9.63	0.798	4.07	0.849
Full BCDA	6.84	9.15	0.824	3.89	0.875

Table 11 Ablation results under different data augmentation strategies for late-stage prediction

Method	MAE	RMSE	R^2	MAPE	ACC±7d
No BCDA	7.28	9.36	0.735	4.14	0.762
Temporal sampling only	5.45	7.08	0.782	3.10	0.833
Full BCDA	4.52	5.89	0.812	2.57	0.881

4.4.2 Ablation on XGBoost constraint settings

This subsection examines the impact of different constraint settings on XGBoost's performance, focusing on whether applying monotonicity constraints enhances prediction accuracy and stability for both early-stage and late-stage tasks. The results of this ablation study are presented in Table 7 and Table 8. Two configurations of Unconstrained XGBoost and Monotone XGBoost were evaluated.

Unconstrained XGBoost shows the weakest performance across the two tasks. In early-stage prediction, it attains an MAE of 7.15 and an RMSE of 9.63 with an R^2 of 0.758. In late-stage prediction, the MAE and RMSE are 4.21 and 5.87, respectively, and R^2 remains at 0.758, suggesting less stable generalization. Introducing monotonicity constraints leads to consistent improvements. Early-stage results improve to MAE 6.84, RMSE 9.15, and R^2 0.824, while late-stage results reach MAE 3.45, RMSE 5.21, and R^2 0.801. These gains indicate that monotonicity constraints help maintain biologically plausible feature effects and improve predictive stability. Overall, the ablation confirms that constraint settings strengthen performance in both tasks, and interaction constraints provide an additional reliability gain when feature relationships are more complex.

4.4.3 Ablation on SC-GAM

This subsection investigates the contribution of *P*-splines and localized curvature control to the performance of the SC-GAM model. The ablation study was conducted by comparing the performance of SC-GAM with *P*-splines and localized curvature control against a configuration without these components. The results of the ablation study are presented in Table 12 and Table 13.

Table 12 Ablation results for early-stage prediction

Method	MAE	RMSE	R^2	MAPE
Without monotone P-splines	7.45	9.87	0.789	4.24
With monotone P-splines	6.18	8.42	0.847	3.52

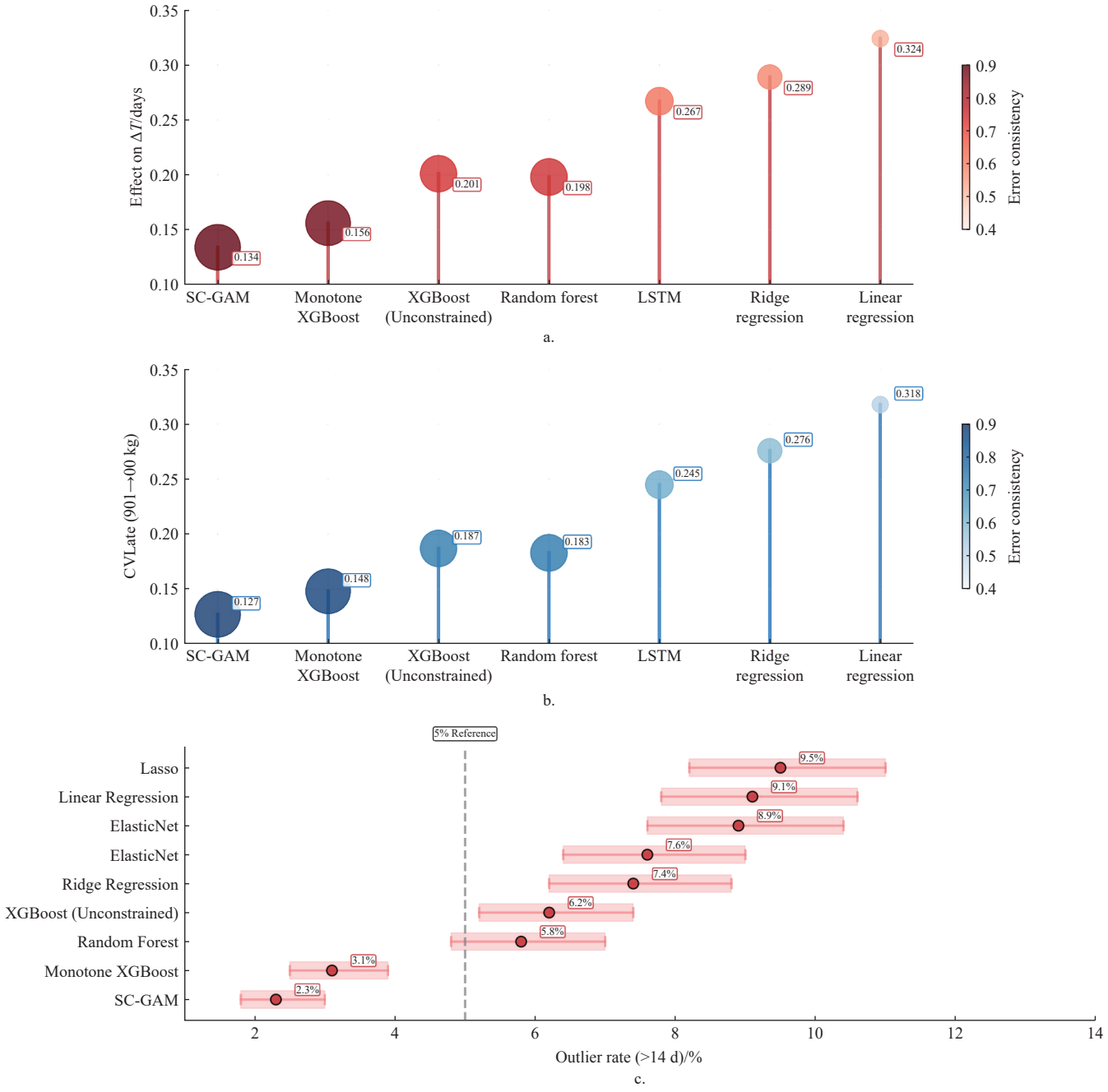
Table 13 Ablation results for late-stage prediction

Method	MAE	RMSE	R^2	MAPE
Without monotone P-splines	4.68	6.21	0.764	2.66
With monotone P-splines	3.45	4.73	0.823	1.96

The ablation results indicate that both *P*-splines and localized curvature control play a critical role in ensuring stable and biologically plausible predictions. Removing these components led to less stable predictions, particularly in the late-stage growth phase.

4.5 Error patterns and robustness

This study assessed model stability beyond point accuracy, focusing on resampling and performance across horizons. Results are summarized in Figure 8 and Table 14.



a. and b. Stability maps for the early-stage and late-stage tasks c. Outlier rate under a ±14 d tolerance with 95% Wilson CIs

Figure 8 Error patterns and robustness across Δ*T* tasks

Table 14 Robustness metrics under grouped five-fold CV

Model	CV (Early)	CV (Late)	Error consistency	Outlier rate
SC-GAM	0.314	0.127	0.892	2.3%
Monotone XGBoost	0.156	0.148	0.876	3.1%
XGBoost (unconstrained)	0.201	0.187	0.739	6.2%
Random forest	0.198	0.183	0.742	5.8%
LSTM	0.267	0.245	0.623	8.9%
Ridge regression	0.289	0.276	0.587	7.4%
Linear regression	0.324	0.318	0.521	9.1%

First, fold-to-fold dispersion was lower for the biology-constrained models in both Δ*T* tasks. Relative to the strongest unconstrained ensemble, the coefficient of variation shrinks by about one third at the early horizon and by a comparable margin at the late horizon, indicating that predictions are less sensitive to which fold a given animal happens to fall into. This aligns with the model design: monotonicity prevents direction reversals when the most recent weight is higher, end-phase tapering dampens volatility around 90-100 kg, and BCDA densifies informative regions without breaking grouped-CV hygiene.

Second, errors align better across horizons with biological

constraints. The error-consistency coefficient, which is Pearson's ρ, is highest for constrained models. This suggests that constrained learners capture stable growth mechanisms that persist when the decision horizon shortens, rather than exploiting horizon-specific idiosyncrasies.

Operationally, extreme misses are rarer under constraints. Under the ±14 d managerial tolerance, the outlier rate drops to a few percent with tight 95% Wilson CIs (Figure 8c), roughly halving the typical rate observed for strong unconstrained baselines. Fewer extreme errors translate directly into fewer premature or delayed market decisions. Taken together, the constrained models are not only more accurate but also more stable and safer to deploy: variability is lower, error structure is more coherent across horizons, and tail risk is reduced.

Lower fold-to-fold CV of RMSE indicates less sensitivity to which animals happen to fall into each fold, which is crucial for small, heterogeneous farms. The high early-late error consistency (ρ) suggests that the same biological signals are being captured across horizons, which explains the stable late-task improvement without retuning.

The outlier-rate analysis under ±14 d is operationally significant, with tail constraints preventing premature shipping and reducing holding costs. Combined with BCDA, constrained models offer a lower-risk operating point with minimal sacrifice in speed or computing.

4.6 Biological realism and constraint compliance

This study assessed whether predictions respect biological regularities using the three checks: the Monotonicity Violation Rate (MVR) within individuals, the Early-Late Consistency Index (ELI) across horizons, and the Non-Negativity Violations (NNV) near the finishing phase. These were aggregated into a composite Biological Score with preset weights.

Table 15 shows a clear pattern. Biology-constrained models dominate across all checks. SC-GAM attains the lowest MVR and zero NNV while achieving the highest ELI, leading to the top Biological Score. Monotone XGBoost ranks second with similarly low violation rates and strong early-late agreement. In contrast, unconstrained ensembles and linear baselines exhibit frequent monotonicity failures and occasional negative outputs around 90-100 kg, which depresses their scores despite sometimes reasonable point accuracy. The unconstrained XGBoost is illustrative: good RMSE, yet double-digit MVR and non-zero NNV indicate biologically unrealistic local oscillations that are corrected once monotonicity and tapering are enforced.

Table 15 Biological plausibility and constraint compliance

Model	MVR/%	ELI	NNV/%	Biological score
SC-GAM	0.8	0.947	0.0	0.973
Monotone XGBoost	1.2	0..923	0.0	0.951
Random forest	8.4	0.834	2.1	0.749
XGBoost (unconstrained)	12.3	0.798	3.7	0.681
LSTM	18.7	0.712	7.2	0.537
Ridge regression	15.6	0.756	4.3	0.628
Linear regression	21.4	0.689	8.9	0.492

These compliance gains are not merely cosmetic. Models with higher Biological Scores also show higher tolerance accuracy and lower outlier rates, implying safer deployment for batch scheduling and market-window decisions. In practical terms, the constrained family reduces the need for post-hoc rule fixing and yields predictions that align with herd-level growth physiology.

Low MVR and NNV reflect respect for basic physiology. The higher the current body weight, the completion time should not increase; the prediction approaching completion should remain non-negative. Their association with ACC±7d and outlier reduction implies that plausibility is not cosmetic: it improves actionable accuracy. This is important for stakeholder trust and reduces the need for rule-based post-filters.

The composite Biological Score highlights a trade-off frontier: models that slightly underperform on point RMSE but excel on MVR, ELI, and NNV can deliver a safer decision boundary. For farms prioritizing risk control over marginal MAE, the constrained models provide a clearly preferable choice.

4.7 Feature importance and predictive insights

Using out-of-fold predictions from the grouped five-fold cross-validation, this study computed the feature importances. These values were standardized within each learner, with the total sum equal to 1, and then averaged across SC-GAM, Random Forest, and monotone XGBoost. The resulting hierarchy is physiologically coherent. Body weight milestone ages are the dominant drivers, contributing about 44% on average. Recent growth-rate descriptors, namely the daily weight gain over intervals and short-term weight gain, also contribute a comparable share of about 43%. Maternal and reproductive records add a smaller but stable signal of about 10%, and breeding or genetic proxies play a minor role of roughly 3%. See Table 16.

Table 16 Top-ranked features for ΔT prediction under grouped five-fold CV

Rank	Feature	Feature type	SC-GAM	Random forest	XGBoost	Average
1	day_30kg	Body weight milestone ages	0.248	0.235	0.241	0.241
2	age_30_100	Growth-rate descriptors	0.186	0.192	0.179	0.186
3	delta_day_30_90	Growth-rate descriptors	0.154	0.148	0.161	0.154
4	day_90kg	Body weight milestone ages	0.127	0.134	0.123	0.128
5	delta_weight_30_90	Growth-rate descriptors	0.089	0.095	0.086	0.090
6	day_70kg	Body weight milestone ages	0.067	0.071	0.074	0.071
7	avg_birth_weight	Reproductive and maternal records	0.045	0.052	0.049	0.049
8	sow_productivity	Reproductive and maternal records	0.032	0.038	0.035	0.035
9	parity	Breeding and genetic records	0.028	0.021	0.031	0.027
10	litter_size_group	Reproductive and maternal records	0.024	0.014	0.022	0.020

To interpret directions of effect rather than only their magnitudes, this study paired the ranking with model-specific explanations. Figure 9 shows the SC-GAM partial shape effect of day_90kg on ΔT with 95% confidence intervals: the effect increases monotonically and then tapers between 90 and 100 kg, consistent with end-phase saturation of growth. Figure 10 summarizes per-sample contributions from the monotone XGBoost model using a SHAP beeswarm. Higher values of day_30kg and day_90kg shift predictions upward, indicating longer ΔT, whereas higher adg_30_100 and recent interval gains from 30-90 d and from 30 to 100 kg shift ΔT downward. Maternal traits, such as average birth weight and sow productivity, contribute modest but consistent adjustments across folds.

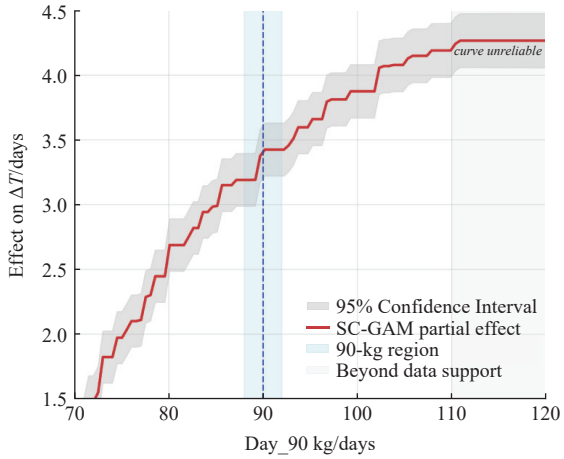


Figure 9 SC-GAM partial effect on ΔT prediction

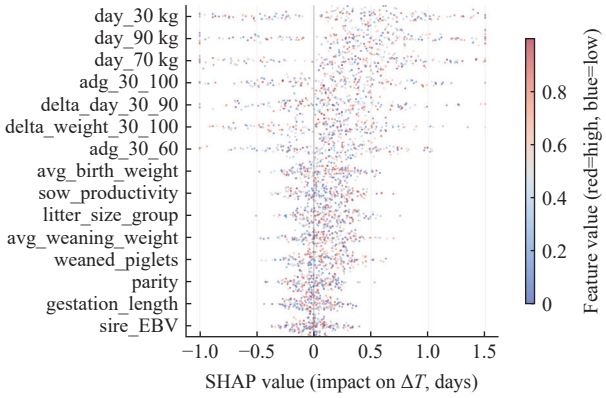


Figure 10 SHAP beeswarm for the monotone XGBoost (top-15 features)

From a management standpoint, these signals are actionable. Milestone ages enable early identification of slow growers, while 30-60 d and 30-100 d ADG supports batch scheduling and market-window decisions. Maternal records refine pen-level management, and genetic proxies add small but fold-stable cues. Taken together, these patterns explain why the biology-constrained learners maintain high accuracy without sacrificing biological plausibility, particularly near 90-100 kg finishing phase.

The importance hierarchy is physiologically coherent. Milestone age is superior to recent ADG, which is superior to maternal characteristics, and in turn superior to genetic indicators. And this explains why constrained learners generalize: monotone effects on milestone ages encode the expectation that animals closer to 100 kg should have a shorter ΔT, while tapering reflects end-phase saturation. SHAP patterns for monotone XGBoost corroborate these mechanisms from an independent model family.

For decision support, the insights are actionable: milestone ages facilitate early screening of slow growers; 30-60 d and 30-100 d ADG support batch scheduling; maternal traits refine pen-level management. This study also noted diminishing returns for purely genetic proxies due to their usefulness as priors but insufficiency alone, thus arguing for continued emphasis on routinely collected phenotypes.

5 Discussion

5.1 Limitations and practical implications for on-farm deployment

This study focuses on predicting time-to-100 kg (ΔT) for finishing pigs using routinely recorded traits from 30 to 100 kg. Extrapolating to lighter growers, alternative terminal weights, or different rations requires recalibrating tapering strength and monotonic signs in the constrained learners and the BCDA pipeline. This aligns with recent PLF reviews that emphasize site-specific configuration when moving from trials to production deployments[25].

Real farms face fragmented data, uneven sensor coverage, and seasonal or feed-formula changes. To prevent leakage and overestimates, grouped five-fold CV by animal was used, following best practices for dependent samples; blocked strategies are recommended when group observations are correlated[14,26]. For post-deployment drift, site-specific fine-tuning in the last 4-6 weeks, along with a drift monitor and periodic refresh of monotone direction, is suggested.

Since shipping and market-window decisions are discrete, tolerance accuracy (ACC±7d) is more actionable than point error. This study recommend pairing ΔT with conformal or quantile bands to obtain distribution-free, coverage-guaranteed intervals that flag pens for re-weighing or diet adjustment; conformal prediction provides finite-sample guarantees and has efficient regression instantiations via conformal quantile regression[15,27]. Figure 11 shows the trade-off between interval width and coverage, with CQR achieving a favorable balance for deployment with narrower intervals and ≥90% coverage.

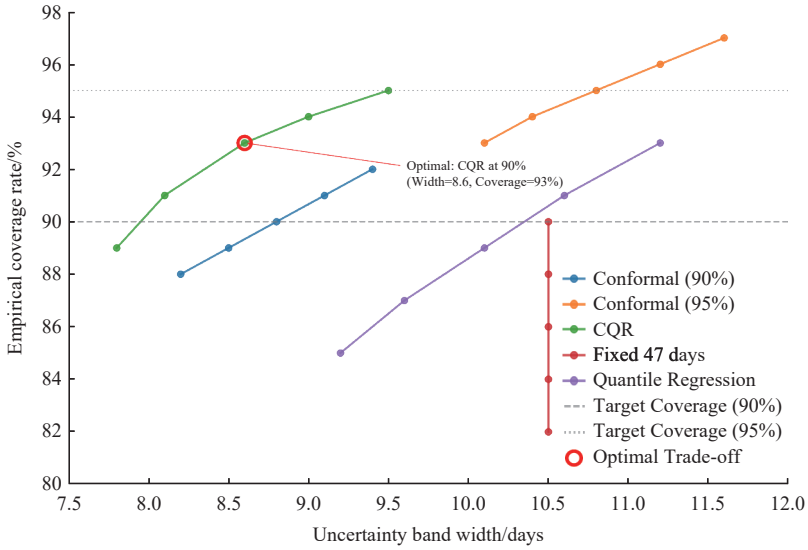


Figure 11 Coverage-width trade-off of uncertainty bands

The learned structure offers practical tools: milestone ages for early screening of slow growers, daily gains from 30-60 kg and 90-100 kg for batch stabilization, and maternal records for pen-level management. A simple cost model can convert ΔT into expected margins per pen, aiding transparent triage and timely shipping.

Constrained models are small and fast, suitable for edge deployment on barn PCs with nightly retraining from the farm database and fixed seeds for reproducibility. Where tree ensembles are preferred, monotone constraints in XGBoost can enforce biologically consistent signs without sacrificing speed.

Three promising avenues include adaptive BCDA for farm-specific bounds, multi-objective tuning for risk-aware optimization, and modern tabular diffusion models for rare regimes. Overall, the biology-constrained family with BCDA offers a lower-risk operating point with minimal latency and computing cost, ready for commercial farm use with routine calibration.

5.2 Engineering implementation and model validation

To further assess the applicability of the model in real-world farm environments, this study applied it to a new batch of pigs and made predictions regarding their growth trajectories.

During the shadow testing phase, this study utilized these models (SC-GAM and Monotone XGBoost) to predict the finishing dates and growth curves for multiple batches of pigs, which were then compared with the actual observed data. Figure 12 illustrates the comparison between the predicted and actual dates. For the early-stage predictions, SC-GAM outperformed Monotone XGBoost. In the late-stage predictions, SC-GAM still showed better performance. The results indicate that SC-GAM outperforms Monotone XGBoost in terms of both predictive accuracy and stability, particularly in forecasting earlier finishing times.

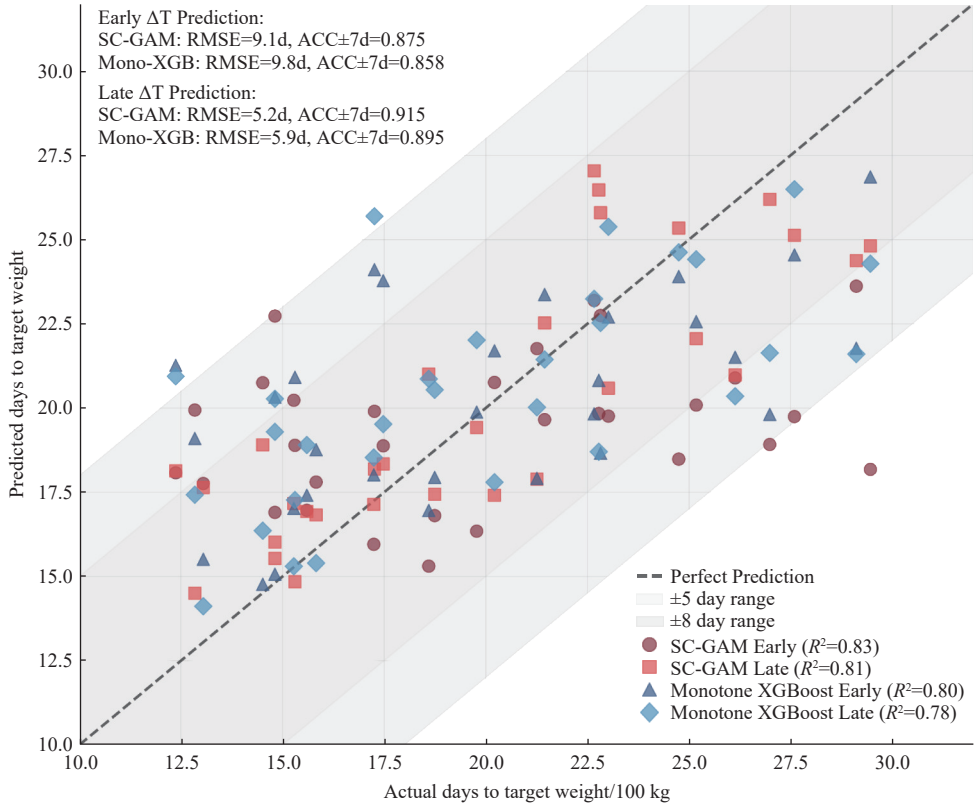


Figure 12 Comparison of predicted and actual days to target weight for early and late predictions

Additionally, Figure 13 presents the growth curve predictions for several pigs. The relationship between weight and age for each pig is depicted as a curve, accompanied by 90% and 95% prediction bands. The comparison of these curves demonstrates the stability of SC-GAM's predictions across individual animals.

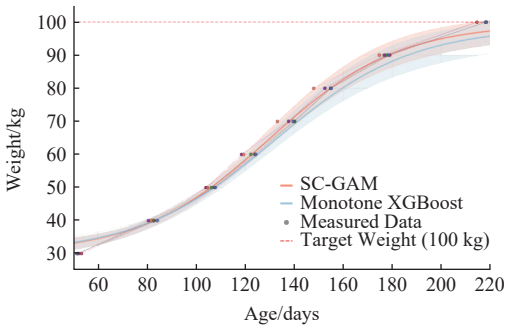


Figure 13 Predicted growth curves for pigs with SC-GAM and Monotone XGBoost models

Figure 14 shows the predicted market window for slaughter, which assists farm managers in optimizing the timing of pig slaughter. By utilizing the predicted market windows, the farm can avoid premature or delayed slaughter decisions, thus enhancing production scheduling, streamlining logistics, and improving overall production efficiency.

Farm managers can use model outputs to monitor each pig's growth and predicted finishing dates in real time, adjusting production plans based on the model's decision support. This shows the model's effectiveness in both experimental and practical farm settings.

The study also compared machine learning models with the classical Gompertz growth model, a widely used biological baseline.

Figure 15 compares predicted growth curves for three pigs from both machine learning models and the Gompertz model. While the Gompertz model works well with sufficient data, it deviates in sparse data conditions, especially near 100 kg. Machine learning

models, however, provide more accurate predictions in both early and late growth stages, even with limited data, highlighting their flexibility and biological consistency.

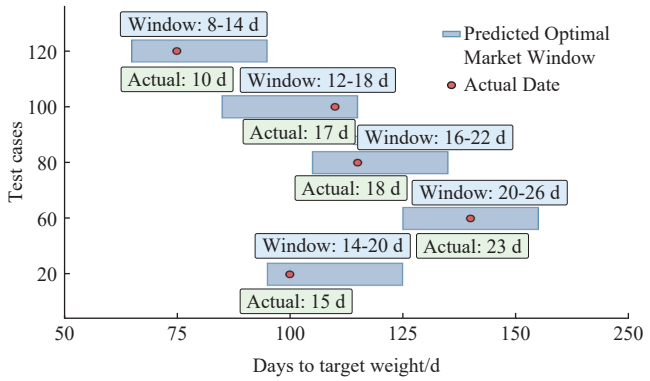


Figure 14 Predicted optimal market windows and actual dates for pigs

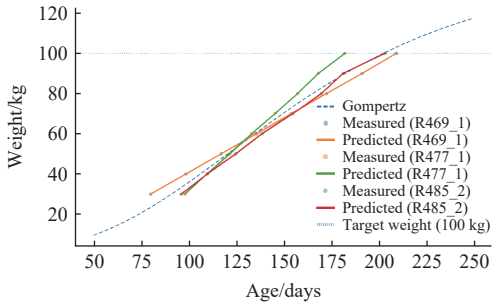


Figure 15 Predicted growth curves of three pigs with a Gompertz reference

5.3 Cross-farm generalization and domain adaptation

A key limitation of this study is that all data were collected from a single commercial farm. While this ensures consistency in recording protocols and data quality, it limits the ability to assess cross-farm generalization. Farm-specific factors, including genetics, feeding strategies, environmental conditions, and management practices, can substantially influence growth trajectories and may affect model performance when applied to new farms.

Despite this limitation, several design choices may improve transferability. First, BCDA framework introduces a strong inductive bias by enforcing universal growth principles, which helps reduce the risk of fitting to farm-specific patterns. Second, the window-based descriptors emphasize growth dynamics within individual animals rather than absolute farm-level benchmarks, potentially mitigating the effects of baseline growth rate differences between farms. Third, the feature-level variance estimation strategy captures within-farm heterogeneity and may improve robustness when transferring to farms with different growth variability profiles.

Future work should prioritize multi-farm validation using rigorous protocols such as leave-one-farm-out evaluation to directly measure cross-farm performance. For deployment under domain shift, practical adaptation strategies could include standardized preprocessing and feature harmonization, covariate-shift reweighting, representation alignment, and lightweight transfer learning with farm-specific fine-tuning and/or domain-adversarial training when a small labeled subset from the target farm is available. Output calibration can further stabilize decision thresholds across farms. Addressing these issues is essential for translating the proposed method into a robust decision-support tool for broader industry applications.

6 Conclusions

This study developed a comprehensive and biologically informed framework for predicting the time-to-100 kg finishing weight in pigs under small-sample, heterogeneous farm conditions. By integrating rigorous preprocessing, BCDA, and biologically constrained learning models, this study addressed long-standing challenges of fragmented records, limited sample sizes, and biologically unrealistic predictions.

The results demonstrate that SC-GAM and monotone XGBoost consistently out-perform unconstrained baselines, not only in point accuracy but also in robustness, tolerance accuracy, and compliance with biological regularities. In particular, the constrained models reduced fold-to-fold volatility, achieved higher early-late consistency, and minimized extreme errors, providing a safer and more interpretable decision-support tool. Feature analyses further confirmed the physiological coherence of the learned structure, highlighting milestone ages and short-term growth rates as primary drivers, with maternal and genetic signals contributing supplementary cues.

Practically, these findings suggest that combining BCDA with biologically constrained models offers a low-risk and computationally efficient operating point for commercial farms. Predictions align with herd growth physiology and support batch scheduling with transparent uncertainty bands.

In summary, this work advances precision livestock farming by showing that biologically constrained machine learning, paired with principled data augmentation and leakage-aware evaluation, can deliver reliable, interpretable, and actionable predictions at farm scale.

Acknowledgements

This work was financially supported by the Key R&D Program of Jiangsu Province (Grant No. BE2023352). The support is gratefully acknowledged. We heartily thank all the experts who took time out of their busy schedules to review this paper and offer their valuable suggestions.

[References]

[1] Trabachini A, Moreira M R, Harada E S, Amorim M N, Silva-Miranda K O. Precision livestock farming applied to swine farms—A systematic literature review. Animals, 2025; 15: 2138.

[2] Marić K. Smart pig farms: Integration and application of digital technologies in pig production. Agriculture, 2025; 15: 937.

[3] Zhou R, Gutzwiller A, Hua J, Weary D M, von Keyserlingk M A G, Tullo E. Data recording and use of data tools for pig health management: A review. Front Vet Sci., 2024; 11: 1490770.

[4] Wang Z, Li Q, Yu Q, Qian W, Gao R, Wang R, et al. A review of visual estimation research on live pig weight. Sensors, 2024; 24: 7093.

[5] Paudel S, de Sousa R V, Sharma S R, Brown-Brandl T. Deep learning models to predict finishing pig weight using point clouds. Animals, 2024; 14: 31.

[6] Kwon K, Park A, Lee H, Mun D. Deep learning-based weight estimation using a fast-reconstructed mesh model from the point cloud of a pig. Comput Electron Agric., 2023; 210: 107903.

[7] Pya N, Wood S N. Shape constrained additive models. Stat Comput, 2015; 25: 543–559.

[8] XGBoost Developers. XGBoost parameters. Available from: https://xgboost.readthedocs.io/en/release_3.0.0/

[9] Hasan M A, Li F, Gouverneur P, Piet A, Grzegorzek M. A comprehensive survey and comparative analysis of time series data augmentation in medical wearable computing. PLoS One, 2025; 20: e0315343.

[10] Iglesias A, Talavera E, González-Prieto A, Mozo A, Gómez-Canaval S. A survey on data augmentation for time series: Taxonomy, directions, and

applications. Neurocomputing, 2023; 558: 126678.
[11] Kotelnikov A, Baranchuk D, Rubachev I, Babenko A. TabDDPM: Modelling tabular data with diffusion models. In: Proceedings of the 40th International Conference on Machine Learning (ICML 2023), PMLR, 2023; Vol. 202. Paper kotelnikov23a.
[12] Sun Q, Zhou X, Zhao S, Gao M, Zhang Q. A comprehensive survey on constrained machine learning: Approaches, models, and applications. Pattern Recognit, 2023; 141: 109529.
[13] Hu Y, Taylor G, Narayan A, Coleman E, Song Y, Sagawa S, et al. Training models that conform to a specification: A practical guide to constrained learning. In: Proceedings of the 40th International Conference on Machine Learning (ICML 2023), PMLR, 2023; pp.14397–14437.
[14] Roberts D R, Bahn V, Ciuti S, Boyce M S, Elith J, Guillera-Arroita G. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography, 2017; 40: 913–929.
[15] Angelopoulos A N, Bates S. A gentle introduction to conformal prediction and distribution-free uncertainty quantification. Found Trends Mach Learn, 2023; 16: 1–140.
[16] Camp Montoro J, Solà-Oriol D, Manteca X, Llonch P, Gasa J, Fàbrega E, et al. Predicting productive performance in grow-finisher pigs using birth and weaning body weight. Animals, 2020; 10: 1017.
[17] Yehoshua A, Bechar A, Cohen Y, Shmuel L, Edan Y. Dynamic sampling algorithm for agriculture-monitoring ground robot. Int J Simul Model, 2023; 22: 392–403.
[18] Cleveland W S. Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc, 1979; 74: 829–836.
[19] Cleveland W S, Devlin S J. Locally weighted regression: An approach to regression analysis by local fitting. J Am Stat Assoc, 1988; 83: 596–610.
[20] Chatfield C. Prediction intervals for the Holt-Winters forecasting procedure. J R Stat Soc Ser C App Stat, 1978; 27: 264–279.
[21] Besteiro J, Góis Á, Madeira M, Manteca-Vilanova R, Rodríguez M R, Fernández M D. Linear and nonlinear mixed models to determine the growth of weaned piglets. Animals, 2023; 13: 2212.
[22] Che Z X, Qiao J K, Xu F J, Li X Y, Zhao Y X, Zhu M J. Integrated analysis reveals genetic basis of growth curve parameters in an F_2 designed pig population based on genome and transcriptome data. Agriculture, 2024; 14(10): 1704.
[23] Henderson C R. Best linear unbiased estimation and prediction under a selection model. Biometrics, 1975; 31: 423–447.
[24] Mrode R A. Linear models for the prediction of animal breeding values. 3rd ed. Wallingford (UK), CABI, 2014.
[25] Akinyemi B E, Vigors B, Turner S P, Akaichi F, Benjamin M, Johnson A K, et al. Precision livestock farming: A qualitative exploration of opportunities and challenges for the commercial pig industry. Front Anim Sci, 2023; 4: 1150528.
[26] Valavi R, Elith J, Lahoz-Monfort J J, Guillera-Arroita G. blockCV: An R package for generating spatially or environmentally separated folds for *k*-fold cross-validation of species distribution models. Methods Ecol Evol, 2019; 10: 225–232.
[27] Romano Y, Patterson E, Candès E. Conformalized quantile regression. Adv Neural Inf Process Syst, 2019; 32: 3543–3553.